



State of Vermont
Agency of Human Services
Department of Mental Health and Mental Retardation
103 South Main Street
Waterbury, Vermont 05671-1601

MEMORANDUM

TO: Performance Indicator Advisory Group

FROM:  John Pandiani and Andy Zovistoski 

DATE: February 18, 1997

SUBJECT: Level of Functioning

The attached graphs and tables provide basic information on level of functioning of CRT clients as measured by QSR reports of annual updates of Global Assessment of Functioning (GAF) Scale scores. For your information, we have also enclosed a copy of the GAF as it appears in DSM IV and an article that discusses a modified GAF scale.

The first graph presents the overall distribution of GAF scores for clients served by each of Vermont's CRT programs during FY 1993 - FY 1996.

The white section at the top of each column represents people for whom no GAF score was reported. As you can see, some programs report more completely than others, and there is a general increase in reporting rates during this time period. Overall, the GAF reporting rates increased from 76% of all CRT clients served during FY 1993, to 89% of all CRT clients served during FY 1996.

The black section at the bottom of each column represents people with a GAF score of 50 or less. The graph refers to this as "severe impairment". The second graph compares the FY 1996 GAF score for people who were assessed at each community mental health center. The proportion of CRT clients who were reported as having a severe impairment varied from 33% at Addison to 70% at Southeast. Overall, 53% of the assessed CRT clients were rated as "severely impaired".

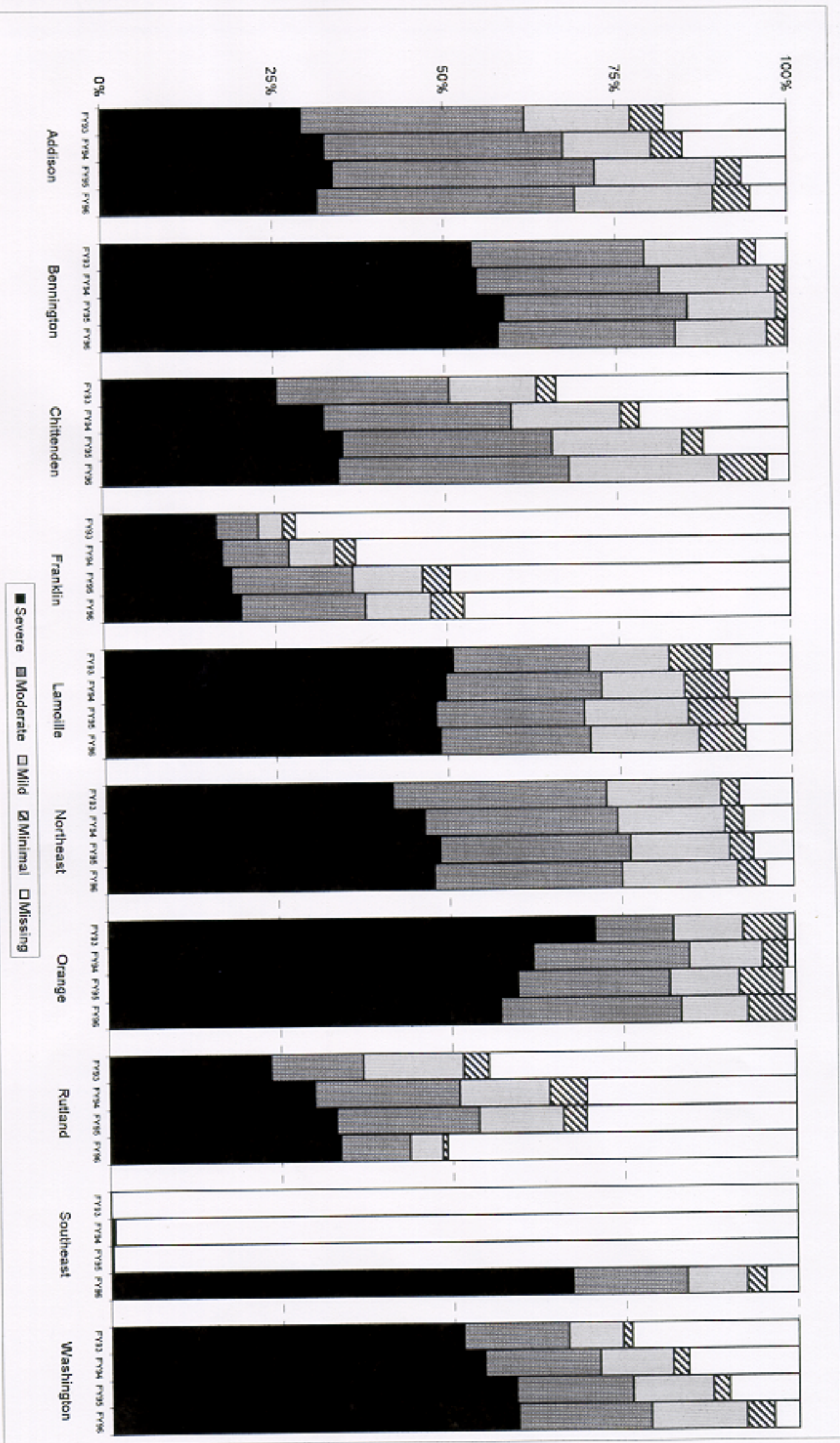
The third graph presents the change in GAF scores over time. The left column in each pair is the difference between 1994 and 1996 scores for the 2,169 people who were served during and have GAF scores for both of those years. The right column in each pair is the difference between 1995 and 1996 scores for the 2,531 people

who were served during and have GAF scores for both of those years. The white area in the center of each column represents people who had identical GAF scores reported for both time periods. For FY 1994-96, there was no change in GAF score reported for 57% of all CRT clients. For FY 1995-96, there was no change in GAF score reported for 71% of all CRT clients.

It occurs to us that the difference between CRT programs may be due to real differences in the level of functioning of people served by these programs, to differences in the assessment practices at the different programs, or to some other factor(s). The preponderance of "no change" over time may be due to real stability in the level of functioning of most clients, or it may be the result of failure to update GAF data reported to the state. If this data is to be useful as a statewide performance indicator, we will need to identify the source(s) of the differences and consistencies observed, and improve on the quality of the data as it is needed.

We will appreciate your thoughts as to the quality of the data, the appropriateness of the analysis, the presentation of the data, and the interpretation of the results.

COMMUNITY REHABILITATION AND TREATMENT PROGRAMS SYMPTOMS/IMPAIRMENT OF CLIENTS SERVED FY1993 - FY1996

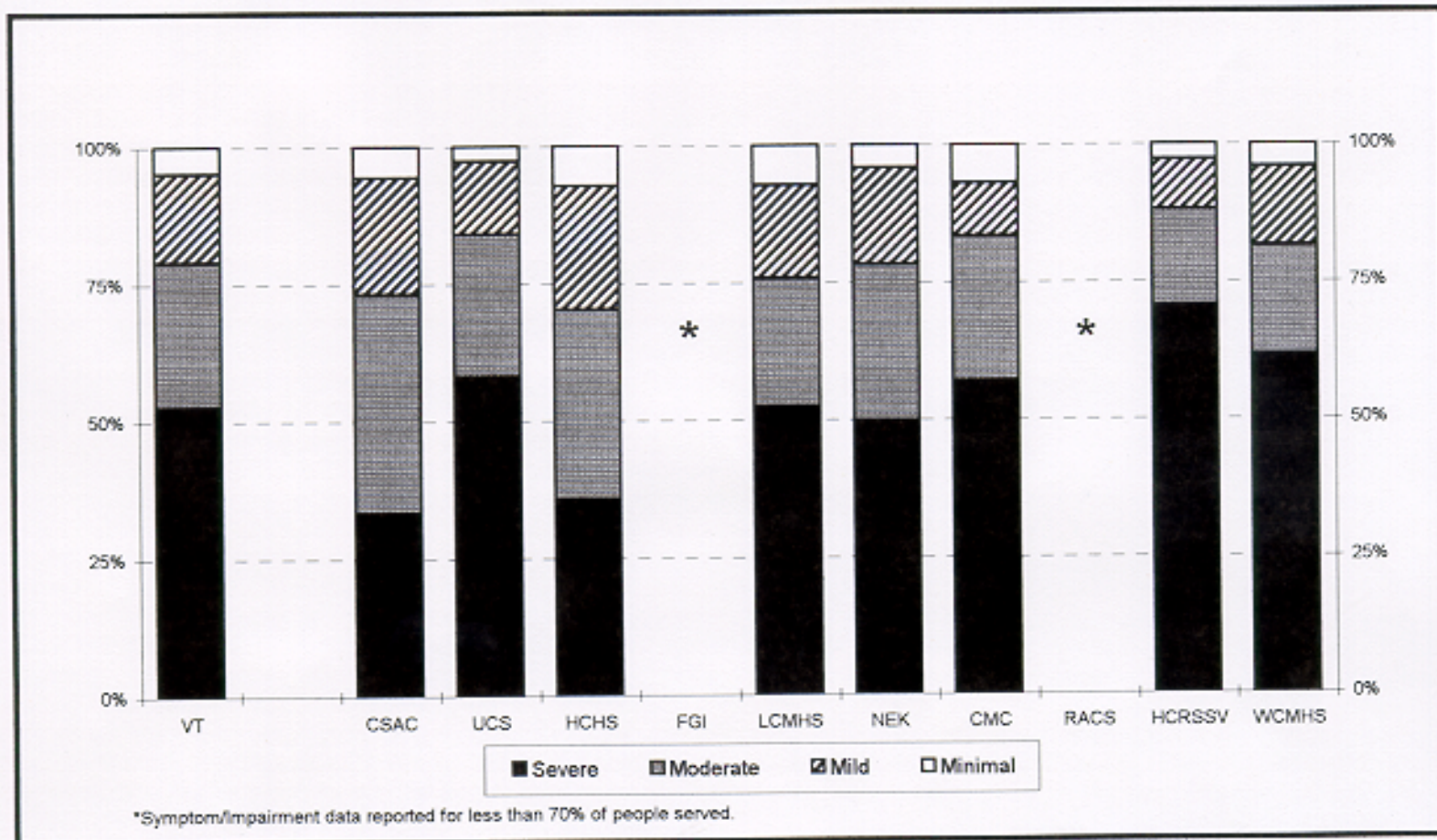


Information is based on Global Assessment of Functioning scale. Ratings are as follows: Severe (1-50); Moderate (51-60); Mild (61-70); and Minimal (71-100).

**COMMUNITY REHABILITATION AND TREATMENT PROGRAMS
SYMPTOMS/IMPAIRMENT OF CLIENTS SERVED FY1993 - FY1996**

Addison CSAC							Bennington UCS															
Severe	Moderate	Mild	Minimal	Missing	Total		Severe	Moderate	Mild	Minimal	Missing	Total										
FY93	36	28%	40	33%	19	0%	6	5%	22	18%	123	116	54%	54	25%	30	14%	5	2%	10	5%	215
FY94	43	33%	46	35%	17	13%	6	5%	20	15%	132	117	55%	57	27%	34	16%	5	2%	1	0%	214
FY95	46	34%	52	38%	24	18%	5	4%	9	7%	136	106	58%	62	27%	30	13%	4	2%	0	0%	222
FY96	47	32%	56	38%	30	20%	8	5%	8	5%	149	130	58%	58	28%	30	13%	6	3%	1	0%	226
Chittenden HCBS							Franklin FGI															
Severe	Moderate	Mild	Minimal	Missing	Total		Severe	Moderate	Mild	Minimal	Missing	Total										
FY93	140	25%	138	25%	70	6%	16	3%	166	34%	550	27	16%	10	6%	6	4%	3	2%	118	72%	164
FY94	189	32%	161	27%	83	16%	16	3%	128	22%	587	34	17%	10	10%	13	7%	6	3%	124	63%	195
FY95	216	35%	184	30%	117	18%	19	3%	77	12%	617	42	19%	40	18%	23	10%	8	4%	112	50%	226
FY96	232	34%	225	33%	147	22%	47	7%	22	3%	673	42	20%	36	16%	20	10%	10	5%	100	48%	210
Lamoille LCMHS							Northeast NEK															
Severe	Moderate	Mild	Minimal	Missing	Total		Severe	Moderate	Mild	Minimal	Missing	Total										
FY93	74	51%	28	20%	17	6%	0	6%	17	12%	146	166	42%	138	31%	74	17%	12	3%	35	5%	440
FY94	70	50%	32	23%	17	12%	0	6%	13	9%	141	223	45%	135	28%	75	16%	13	3%	35	7%	441
FY95	87	48%	30	22%	21	15%	10	7%	11	8%	138	230	48%	136	28%	71	14%	17	3%	29	6%	402
FY96	65	48%	28	22%	21	16%	9	7%	0	7%	133	217	48%	124	27%	77	17%	18	4%	19	4%	455
Orange CMC							Rutland RACS															
Severe	Moderate	Mild	Minimal	Missing	Total		Severe	Moderate	Mild	Minimal	Missing	Total										
FY93	56	71%	9	11%	0	6%	5	6%	1	1%	79	63	24%	36	13%	38	15%	10	4%	120	45%	268
FY94	52	62%	18	23%	9	11%	3	4%	1	1%	84	62	30%	58	21%	36	13%	15	5%	84	31%	275
FY95	65	60%	24	22%	11	10%	7	6%	2	2%	109	87	33%	61	21%	36	12%	10	3%	90	31%	284
FY96	65	57%	30	26%	11	10%	8	7%	0	0%	114	106	34%	32	10%	15	5%	2	1%	161	51%	316
Southeast HCRSSV							Washington WCMHS															
Severe	Moderate	Mild	Minimal	Missing	Total		Severe	Moderate	Mild	Minimal	Missing	Total										
FY93	0	0%	0	0%	0	6%	0	0%	341	100%	341	274	51%	82	15%	42	8%	7	1%	130	24%	535
FY94	1	0%	0	0%	1	0%	0	0%	355	99%	357	335	54%	104	17%	66	11%	14	2%	100	16%	622
FY95	1	0%	0	0%	0	0%	0	0%	358	100%	358	352	58%	102	17%	70	12%	14	2%	81	10%	559
FY96	270	67%	67	17%	35	8%	11	3%	19	5%	402	371	59%	121	19%	87	14%	25	4%	23	4%	627

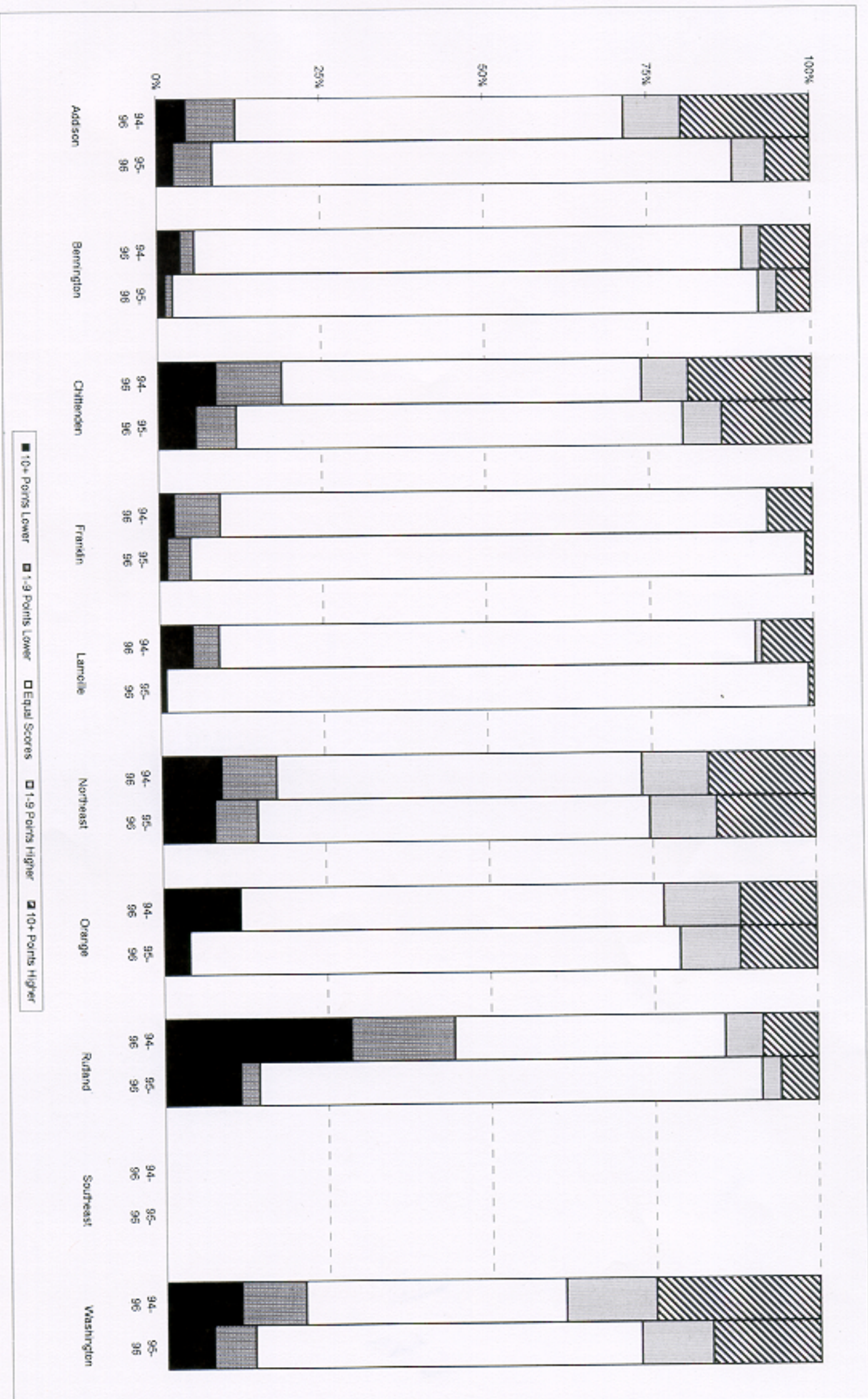
COMMUNITY REHABILITATION AND TREATMENT PROGRAMS SYMPTOMS/IMPAIRMENT OF PEOPLE SERVED, FY1996



Region/Provider	Symptoms / Impairment								Total Reported	Percent with Impairment Data
	Severe		Moderate		Mild		Minimal			
	#	%	#	%	#	%	#	%		
Addison - CSAC	47	33%	56	40%	30	21%	8	6%	141	97%
Bennington - UCS	130	58%	58	26%	30	13%	6	3%	224	100%
Chittenden - HCHS	232	36%	225	35%	147	23%	47	7%	651	97%
Franklin - FGI	42	38%	38	35%	20	18%	10	9%	110	52%
Lamoille - LCMHS	65	52%	29	23%	21	17%	9	7%	124	93%
Northeast -NEK	217	50%	124	28%	77	18%	18	4%	436	96%
Orange - CMC	65	57%	30	26%	11	10%	8	7%	114	100%
Rutland -RACS	106	68%	32	21%	15	10%	2	1%	155	49%
Southeast - HCRSSV	270	70%	67	17%	35	9%	11	3%	383	95%
Washington - WCMHS	371	61%	121	20%	87	14%	25	4%	604	96%
TOTAL	1545	53%	780	27%	473	16%	144	5%	2,942	89%

Information is based on Global Assessment of Functioning scale reported in quarterly service reports submitted by Vermont's community service providers and includes people assigned to Community Rehabilitation and Treatment Programs. Ratings are as follows: Severe (1 - 50); Moderate (51 - 60); Mild (61 - 70); and Minimal (71 - 100).

COMMUNITY REHABILITATION AND TREATMENT PROGRAMS
CHANGE OVER TIME IN SYMPTOMS/IMPAIRMENT: FY1994 - 96 AND FY1995 - 96



**COMMUNITY REHABILITATION AND TREATMENT PROGRAMS
CHANGE IN SYMPTOMS/IMPAIRMENT FOR CLIENTS SERVED IN BOTH FY1994 AND FY1995**

Change in Symptoms/Impairment	Addison CSAC	Bennington UCS	Chittenden HCHS	Franklin FGI	Lamoille LCMHS	Northeast NEK	Orange CMC	Rutland RACS	Southeast HCRSSV	Washington WCMHS	Statewide Total
10+ Points Lower	4	5	31	1	5	25	6	20	0	38	135
7-9 Points Lower	7	3	35	3	4	23	0	11	0	33	118
Equal Scores	54	119	182	36	82	153	32	29	0	133	801
1-6 Points Higher	8	4	25	0	1	28	6	4	0	45	122
10+ Points Higher	18	11	68	3	8	45	6	6	0	84	247
Missing Data	15	0	121	86	11	22	0	148	239	63	715
Total Served Both Years	106	142	470	139	111	296	51	218	239	387	2,109

Information is based on difference in GAF between FY1994 and FY1995.

**COMMUNITY REHABILITATION AND TREATMENT PROGRAMS
CHANGE IN SYMPTOMS/IMPAIRMENT FOR CLIENTS SERVED IN BOTH FY1996 AND FY1995**

Change in Symptoms/Impairment	Addison CSAC	Bennington UCS	Chittenden HCHS	Franklin FGI	Lamoille LCMHS	Northeast NEK	Orange CMC	Rutland RACS	Southeast HCRSSV	Washington WCMHS	Statewide Total
10+ Points Lower	3	2	27	1	1	26	3	12	0	28	103
7-9 Points Lower	7	2	29	3	0	21	0	3	0	25	90
Equal Scores	93	155	321	79	113	193	57	82	0	232	1020
1-6 Points Higher	6	5	28	0	0	33	7	3	0	43	125
10+ Points Higher	8	9	65	1	1	48	9	6	0	65	213
Missing Data	6	0	76	100	7	20	0	150	280	35	677
Total Served Both Years	123	173	546	194	122	345	76	254	280	428	2,501

Information is based on difference in GAF between FY1995 and FY1996.

Global Assessment of Functioning (GAF) Scale

Consider psychological, social, and occupational functioning on a hypothetical continuum of mental health-illness. Do not include impairment in functioning due to physical (or environmental) limitations.

Code (Note: Use intermediate codes when appropriate, e.g., 45, 68, 72.)

- 100 Superior functioning in a wide range of activities, life's problems never seem to get out of hand, is sought out by others because of his or her many positive qualities. No symptoms.
- 90 Absent or minimal symptoms (e.g., mild anxiety before an exam), good functioning in all areas, interested and involved in a wide range of activities, socially effective, generally satisfied with life, no more than everyday problems or concerns (e.g., an occasional argument with family members).
- 80 If symptoms are present, they are transient and expectable reactions to psychosocial stressors (e.g., difficulty concentrating after family argument); no more than slight impairment in social, occupational, or school functioning (e.g., temporarily falling behind in schoolwork).
- 70 Some mild symptoms (e.g., depressed mood and mild insomnia) OR some difficulty in social, occupational, or school functioning (e.g., occasional truancy, or theft within the household), but generally functioning pretty well, has some meaningful interpersonal relationships.
- 60 Moderate symptoms (e.g., flat affect and circumstantial speech, occasional panic attacks) OR moderate difficulty in social, occupational, or school functioning (e.g., few friends, conflicts with peers or co-workers).
- 50 Serious symptoms (e.g., suicidal ideation, severe obsessional rituals, frequent shoplifting) OR any serious impairment in social, occupational, or school functioning (e.g., no friends, unable to keep a job).
- 40 Some impairment in reality testing or communication (e.g., speech is at times illogical, obscure, or irrelevant) OR major impairment in several areas, such as work or school, family relations, judgment, thinking, or mood (e.g., depressed man avoids friends, neglects family, and is unable to work; child frequently beats up younger children, is defiant at home, and is failing at school).
- 30 Behavior is considerably influenced by delusions or hallucinations OR serious impairment in communication or judgment (e.g., sometimes incoherent, acts grossly inappropriately, suicidal preoccupation) OR inability to function in almost all areas (e.g., stays in bed all day; no job, home, or friends).
- 20 Some danger of hurting self or others (e.g., suicide attempts without clear expectation of death; frequently violent; manic excitement) OR occasionally fails to maintain minimal personal hygiene (e.g., smears feces) OR gross impairment in communication (e.g., largely incoherent or mute).
- 10 Persistent danger of severely hurting self or others (e.g., recurrent violence) OR persistent inability to maintain minimal personal hygiene OR serious suicidal act with clear expectation of death.
- 0 Inadequate information.

The rating of overall psychological functioning on a scale of 0–100 was operationalized by Luborsky in the Health-Sickness Rating Scale (Luborsky L: "Clinicians' Judgments of Mental Health." *Archives of General Psychiatry* 7:407–417, 1962). Spitzer and colleagues developed a revision of the Health-Sickness Rating Scale called the Global Assessment Scale (GAS) (Endicott J, Spitzer RL, Fleiss JL, Cohen J: "The Global Assessment Scale: A Procedure for Measuring Overall Severity of Psychiatric Disturbance." *Archives of General Psychiatry* 33:766–771, 1976). A modified version of the GAS was included in DSM-III-R as the Global Assessment of Functioning (GAF) Scale.

Global Assessment of Functioning

A Modified Scale

RICHARD C. W. HALL, M.D.

The modified Global Assessment of Functioning (GAF) scale has more detailed criteria and a more structured scoring system than the original GAF. The two scales were compared for reliability and validity. Raters who had different training levels assigned hospital admission and discharge GAF scores from patient charts. Intraclass correlation coefficients for admission GAF scores were higher for raters who used the modified GAF (0.81), compared with raters who used the original GAF (0.62). Validity studies showed a high correlation (0.80) between the two sets of scores. The modified GAF also correlated well with Zung Depression scores (-0.73). The modified GAF may be particularly useful when interrater reliability needs to be maximum and/or when persons with varying skills and employment backgrounds—and without much GAF training—must rate patients. Because of the increased structure, the modified GAF may also be more resistant to rater bias. (Psychosomatics 1995; 36:267-275)

Global severity of illness scales are important instruments for assessing change in psychiatric patients. Increasingly, such scales are being used by managed care companies and governmental agencies to determine who can and cannot be admitted to hospitals.^{1,2} The scales are simple to administer and are more sensitive to differential treatment effects than measures of single dimensions of psychopathology.^{3,4} Probably the most often used global assessment instrument is the interviewer-rated Global Assessment of Functioning (GAF) scale, which is listed in the DSM-III-R as an Axis V diagnostic criterion test. This scale is very similar to the Global Assessment Scale (GAS) developed by Endicott et al.⁴ The Endicott scale has values that range from 1, representing the sickest patient, to 100, a person with no symptoms. The scale is divided into 10 equal intervals, with 10 scores in each interval, and the criteria that define each score in each interval

are listed. The GAF scale has similar criteria and the same interval design, except that the value range is from 1 to 90 (absent or minimal symptoms), and there are 9 rather than 10 equal intervals.

One potential problem with both these interviewer-rated scales is that for the scores to be comparable and thus meaningful across different studies, interrater reliability in scoring must be quite consistent within a study and from one study to another. Interrater reliability is strongly influenced by two factors: 1) the consistency of the raters, and 2) the heterogeneity of patient

Received March 3, 1993; revised April 26, 1993; accepted May 21, 1993. From the Center for Psychiatry, Florida Hospital, Orlando, Florida, and Department of Psychiatry, University of Florida, Gainesville. Address reprint requests to Dr. Hall, Center for Psychiatry, Florida Hospital, 601 E. Rollins St., Orlando, FL 32803.

Copyright © 1995 The Academy of Psychosomatic Medicine.

Modified GAF Scale

illness severity. Endicott et al.⁴ tested the reliability of the GAS in 5 studies and reported intraclass correlation coefficients ranging from 0.61 to 0.91, with associated standard error of measurement scores ranging from 5.0 to 8.0 units. Most of Endicott's ratings were done by only a few, well-trained interviewers. Having consistently trained interviewers should produce a greater likelihood of higher interrater reliability scores and smaller standard errors. Yet even with this bias, two of the studies had intraclass correlation coefficients in the 0.60s, suggesting that the scale might be less reliable than had been hoped. In contrast, one of the reliability studies used 15 raters of different backgrounds and training levels. Although the intraclass correlation coefficient was high, this was due primarily to a greater heterogeneity of illness severity as compared to the other studies, not to interrater consistency of scoring. The lack of interrater consistency was demonstrated by a high standard error of measurement not seen in the other studies.

Although we could not find published reliability studies on the GAF in the literature, our subjective experience at Florida Hospital was that the GAF was used by staff members of different backgrounds (physicians with varying degrees of familiarity with the scale, nurses, Ph.D. researchers), and GAF ratings from these staff differed substantially for the same patient. Thus, we hypothesized that the original GAF might be less reliable than we had expected. To test this hypothesis and to improve interrater reliability, we developed a modified GAF scale, and we formally tested interrater reliability in the original and modified versions of the GAF. We conducted our study in 1992-1993.

METHODS

A modified GAF scale was developed by increasing the structure of the original GAF instrument with a greater number of criteria and with additional directions for assigning scores. We chose to modify the GAF rather than the GAS because the GAF, as listed in the DSM-III-R, reflects more current ideas on illness se-

verity rating and is the more frequently used instrument. The criteria and scoring changes that we made in the GAF were tested among a small group of staff members who rated patients from successive drafts of the modified scale. When staff members had different ratings of a given patient, their reasons were discussed, and changes were made in the wording or use of the criteria or scoring directions.

Reliability studies for both the original and modified GAF scales were based on ratings of 16 patient intake histories and discharge summaries taken from the patients' hospital charts. All of these patients had diagnoses of major depression with or without comorbid eating disorders. They had all been inpatients on the Affective/Eating Disorders Unit, and their intake histories were obtained by one of the same two doctors. These particular 16 patients were chosen for review because they had the most detailed intake histories and discharge summaries available. Thus, a maximum amount of patient information was available for evaluation with the GAF.

Two groups of staff from the psychiatric units at Florida Hospital rated each of the same 16 patient histories and discharge summaries. All patients were given a GAF score for the severity of illness at admission and a second GAF score for illness severity at discharge. One group of staff rated the patients using the original GAF, and the other group of staff rated the patients using the modified GAF. None of the staff received any training in the use of either GAF, but they were allowed to read it and to ask questions for clarification. This procedure was followed to evaluate the consistency of ratings by untrained staff; therefore, we could evaluate the soundness and reliability of each GAF under these conditions.

The staff in the group using the original GAF consisted of 12 professionals (nurses, physicians, social workers, psychiatry technicians, and clinical Ph.D.'s) assigned to 2 inpatient treatment units (affective/eating disorders and psychiatric/medical). The staff in the group using the modified GAF consisted of another group of 12 professionals from other inpatient

units (acute general psychiatry, adolescent, or intensive treatment). Within each of the rating groups (original or modified GAF), the means and standard errors were calculated for the ratings of each patient on admission and discharge. Intraclass correlation coefficients (ICC) were then calculated separately for the original GAF group on admission and discharge and for the modified GAF group on admission and discharge. Both the admission and discharge correlation coefficients were compared between the groups.

The concurrent validity of the modified GAF was tested by comparing admission scores of this instrument with admission scores on the original GAF, the Zung depression test, and a self-rating of global illness severity. Pearson Product Moment correlations were used for these three assessments of validity. For the modified and original GAF comparison, admission scores were obtained from the same 16 patient histories and discharge summaries as in the reliability tests. For the modified GAF and Zung comparison and the modified GAF and self-rating of illness comparison, data were obtained from outpatient telephone interviews with 142 patients who had been discharged from Florida Hospital 6 months to 1.5 years before. These patients all had diagnoses of major depression with or without comorbid diagnoses of eating disorder. Each patient had been evaluated using the modified GAF only, the Zung depression test, and a self-illness severity rating. The self-rated global illness scores were on a scale of 1-10, where 1 was sickest and 10 was most healthy.

RESULTS

The Modified GAF

The modified GAF retained the same 1-90 scale with the same 10-point intervals as the original GAF. All criteria in the original GAF were retained and were listed on separate lines to facilitate quick reading (Table 1). Additional criteria were added to most of the 10-point intervals, and directions for scoring the pa-

tient's illness severity were added at the end of each 10-point interval. The purpose for these additions was to decrease the variability in scoring. Usually, the scoring within a 10-point interval applied only to the criteria within that interval. For example, in the 81-90 interval, a patient having no symptoms or problems received a score of 88-90; a patient having minimal symptoms or problems received a score of 84-87; and a patient having minimal symptoms and problems received a score of 81-83 (Table 1). However, in the 21-30, 31-40, and 41-50 scoring intervals, the same 10 criteria were listed in each interval, and the score depended on the number of criteria that a patient met within these 3 scoring intervals.

For example, if a patient met 1 of these criteria, the score was 48-50; if a patient met 2 of the criteria, the score was 44-47; and if the patient met 3 of the criteria, the score was 41-43. However, if the patient met 4-6 of the criteria, the scores ranged from 31-40. If the patient met 7-10 of the criteria, the scores ranged from 21-30 (Table 1). Finally, in the 21-30 scoring interval, a unique set of criteria and scores also existed in addition to the criteria and scoring already discussed. These unique criteria were listed in the original GAF and were deemed to be of sufficient seriousness that they should not be added to the list of criteria in the 31-40 and 41-50 intervals but rather would warrant the lowest score available in the 21-30 category. Thus, suicidal preoccupation and preparation, behavior considerably influenced by delusions or hallucinations, or serious impairment in communication (i.e., sometimes incoherent or profound stuporous depression), always elicited a score of 21.

The various changes we made in modifying the GAF made it longer than the original GAF (4 pages vs. 1). Thus, it is suggested that when using this new GAF, the interviewer should question the patient about each of the criteria, then write down answers, and later count the number of criteria that the patient meets. It is felt that the slower speed in assigning a score from the modified GAF is compensated for by the increased consistency of ratings attributable

TABLE 1. Modified Global Assessment of Functioning (GAF) scale

Guidelines for Scale Use:

1. One Admission GAF score, evaluated on history and symptoms at admission
One Discharge GAF score, evaluated on symptoms at discharge
One Outpatient GAF score, evaluated on symptoms during current past month
2. Rate LOWEST possible score for each patient

Score
Intervals

90	Absent or Minimal Symptoms	Criteria: Minimal or absent symptoms (e.g., mild anxiety before an examination) Good functioning in all areas and satisfied with life Interested and involved in a wide range of activities Socially effective No more than everyday problems or concerns (e.g., an occasional argument with family members)	Scoring: A patient with no symptoms or everyday problems = rating 88-90 A patient with minimal symptoms or everyday problems = rating 84-87 A patient with minimal symptoms and everyday problems = rating 81-83
80	Some Transient Mild Symptoms	Criteria: Mild symptoms are present, but they are transient and expectable reactions to psychosocial stressors (e.g., difficulty concentrating after family argument) Slight impairment in social, work, or school functioning (e.g., temporarily falling behind in school or work)	Scoring: A patient with EITHER mild symptom(s) OR mild impairment in social, work, or school functioning = rating 78-80 A patient with mild impairment in more than 1 area of social, work, or school functioning = rating 74-77 A patient with BOTH mild symptoms AND slight impairment in social, work, and school functioning = rating 71-73
70	Some Persistent Mild Symptoms	Criteria: Mild symptoms are present that are NOT just expectable reactions to psychosocial stressors (e.g., mild or lessened depression and/or mild insomnia) Some persistent difficulty in social, occupational, or school functioning (e.g., occasional truancy, theft within the family, or repeated falling behind in school or work) But has some meaningful interpersonal relationships	Scoring: A patient with EITHER mild persistent symptoms OR mild difficulty in social, work, or school functioning = rating 68-70 A patient with mild persistent difficulty in more than 1 area of social, work, or school functioning = rating 64-67 A patient with BOTH mild persistent symptoms AND some difficulty in social, work, and school functioning = rating 61-63

60	Moderate Symptoms	Moderate symptoms (e.g., frequent, moderate depressed mood and insomnia and/or moderate ruminating and obsessing; or occasional anxiety attacks; or flat affect and circumstantial speech; or eating problems and below minimum safe weight without depression)
	Criteria:	Moderate difficulty in social, work, or school functioning (e.g., few friends or conflicts with co-workers)
	Scoring:	A patient with EITHER moderate symptoms OR moderate difficulty in social, work, or school functioning = rating 58-60
51		A patient with moderate difficulty in more than 1 area of social, work, or school functioning = rating 54-57
		A patient with BOTH moderate symptoms AND moderate difficulty in social, work, and school functioning = rating 51-53
50	Some Serious Symptoms or Impairment in Functioning	Serious impairment with work, school, or housework if a housewife or househusband (e.g., unable to keep job or stay in school, or failing school, or unable to care for family and house)
	Criteria:	Frequent problems with the law (e.g., frequent shoplifting, arrests) or occasional combative behavior
		Serious impairment in relationships with friends (e.g., very few or no friends, or avoids what friends he/she has)
		Serious impairment in relationships with family (e.g., frequent fights with family and/or neglects family or has no home)
		Serious impairment in judgment (including inability to make decisions, confusion, disorientation)
		Serious impairment in thinking (including constant preoccupation with thoughts, distorted body image, paranoia)
		Serious impairment in mood (including constant depressed mood plus helplessness and hopelessness, or agitation, or manic mood)
		Serious impairment due to anxiety (panic attacks, overwhelming anxiety)
		Other symptoms: some hallucinations, delusions, or severe obsessional rituals
		Passive suicidal ideation
	Scoring:	A patient with 1 area of disturbance = rating 48-50
41		A patient with 2 areas of disturbance = rating 44-47
		A patient with 3 areas of disturbance = rating 41-43
40	Major Impairment in Several Areas of Functioning	Serious impairment with work, school, or housework if a housewife or househusband (e.g., unable to keep job or stay in school, or failing school, or unable to care for family and house)
	Criteria:	Frequent problems with the law (e.g., frequent shoplifting, arrests) or occasional combative behavior
		Serious impairment in relationships with friends (e.g., very few or no friends, or avoids what friends he/she has)
		Serious impairment in relationships with family (e.g., frequent fights with family and/or neglects family or has no home)
		Serious impairment in judgment (including inability to make decisions, confusion, disorientation)
		Serious impairment in thinking (including constant preoccupation with thoughts, distorted body image, paranoia)
		Serious impairment in mood (including constant depressed mood plus helplessness and hopelessness, or agitation, or manic mood)
		Serious impairment due to anxiety (panic attacks, overwhelming anxiety)
		Other symptoms: some hallucinations, delusions, or severe obsessional rituals
		Passive suicidal ideation
	Scoring:	A patient with 4 areas of disturbance = rating 38-40
31		A patient with 5 areas of disturbance = rating 34-37
		A patient with 6 areas of disturbance = rating 31-33

Modified GAF Scale

TABLE 1. Modified Global Assessment of Functioning (GAF) scale (continued)

Score Intervals	
30	Inability to Function in Almost All Areas Criteria: Unique Items: OR Combined Items: Suicidal preoccupation or frank suicidal ideation with preparation OR behavior considerably influenced by delusions or hallucinations OR serious impairment in communication (sometimes incoherent, acts grossly inappropriately, or profound stuporous depression) OR serious impairment with work, school, or housework if a housewife or househusband (unable to keep job or stay in school, or failing school, or unable to care for family and house) Frequent problems with the law (frequent shoplifting, arrests) or occasional combative behavior Serious impairment in relationships with friends (very few or no friends, or avoids what friends he/she has) Serious impairment in relationships with family (frequent fights with family and/or neglects family or has no home) Serious impairment in judgment (including inability to make decisions, confusion, disorientation) Serious impairment in thinking (including constant preoccupation with thoughts, distorted body image, paranoia) Serious impairment in mood (including constant depressed mood plus helplessness and hopelessness, or agitation, or manic mood) Serious impairment due to anxiety (panic attacks, overwhelming anxiety) Other symptoms: some hallucinations, delusions, or severe obsessional rituals Passive suicidal ideation A patient with 1 of the first 3 (unique) criteria = rating 21 OR a patient with 7 of the combined criteria = rating 28-30 A patient with 8-9 of the combined criteria = rating 24-27 A patient with 10 of the combined criteria = rating 20-23
20	In Some Danger of Hurting Self or Others Criteria: Suicide attempts without clear expectation of death (e.g., mild overdose or scratching wrists with people around) Some severe violence or self-mutilating behaviors Severe manic excitement, or severe agitation and impulsivity Occasionally fails to maintain minimal personal hygiene (e.g., diarrhea due to laxatives, or smearing feces) Urgent/emergency admission to the present psychiatric hospital In physical danger due to medical problems (e.g., severe anorexia or bulimia and some spontaneous vomiting or extensive laxative/diuretic/diet pill use, but without serious heart or kidney problems or severe dehydration and disorientation) A patient with 1-2 of the 6 areas of disturbance in this category = rating 18-20 A patient with 3-4 of the 6 areas of disturbance in this category = rating 14-17 A patient with 5-6 of the 6 areas of disturbance in this category = rating 11-13

10	In Persistent Danger of Severely Hurting Self or Others	Criteria:	Serious suicidal act with clear expectation of death (e.g., stabbing, shooting, hanging, or serious overdose, with no one present) Frequent severe violence or self-mutilation Extreme manic excitement, or extreme agitation and impulsivity (e.g., wild screaming and ripping the stuffing out of a bed mattress) Persistent inability to maintain minimal personal hygiene Urgent/emergency admission to present psychiatric hospital In acute, severe danger due to medical problems (e.g., severe anorexia or bulimia with heart/kidney problems, or spontaneous vomiting WHENEVER food is ingested, or severe depression with out-of-control diabetes)	Scoring:	A patient having 1-2 of the 6 areas of disturbance in this category = rating of 8-10 A patient having 3-4 of the 6 areas of disturbance in this category = rating of 4-7 A patient with 5-6 of the 6 areas of disturbance in this category = rating of 1-3	0
----	---	-----------	---	----------	--	---

to the increased detail and structure in the instrument.

Reliability of Original and Modified GAF Scales

Both for admission and for discharge the intraclass correlation coefficients were higher in the modified GAF group as compared with the original group. The difference in ICC was particularly striking on admission, where it was 0.62 for the original group and 0.81 for the modified group. At discharge the ICC for the original group was 0.90 and for the modified group, 0.95 (Table 2).

Because the original and modified GAF groups used the same patient histories, the source data were comparable. Thus, the difference in the admission ICC's between groups would appear to be primarily due to a greater interrater variability rather than patient heterogeneity. Support for this idea can be found in the standard errors of the ratings for each patient, which reflect interrater variability. Of the 16 standard errors for the patient's admission data, 13 were higher in the original GAF group than in the modified group. Thus, there was more variability among rater's GAF scores in the group using the original GAF.

TABLE 2. Global Assessment of Functioning (GAF) scale and psychiatric tests: correlation coefficients

	Intraclass Correlation Coefficients	
	Admission GAF Scores	Discharge GAF Scores
Original GAF test	0.62	0.90
Modified GAF test	0.81	0.95
	Pearson Product Moment Correlations	
	All Outpatient Scores	
Modified GAF and Zung depression test	-0.73	
Modified GAF and self-rated illness severity scale	0.58	
	Admission GAF Scores	
Modified GAF and original GAF	0.80	

Modified GAF Scale

Interestingly, all of the means for the patient's admission GAF scores were also higher in the original GAF group than in the modified group. Thus, the modified GAF caused patients to be rated more sick than the original GAF.

Concurrent Validity

Because all of the mean admission GAF scores for the original group were higher than the scores in the modified group, we wanted to test the correlation between the scores of the two GAF's and test the correlation of the modified GAF with other psychological assessment tests. The Pearson Product Moment correlation coefficient between the 16 original and 16 modified mean admission scores was 0.80, $P < 0.001$, $df = 14$, showing good correlation (Table 2).

Because all of the patients used in these studies were depressed, we also compared modified GAF scores with the scores from the Zung depression test. The Pearson Product Moment correlation coefficient was -0.73, $P < 0.001$ (negative because a higher number represents sickness in the Zung scores and a lower number represents sickness in the GAF) (Table 2).

Finally, we also correlated modified GAF scores with the scores that patients gave themselves to indicate their severity of illness. The Pearson Product Moment correlation coefficient was 0.58, $P < 0.01$ (Table 2).

DISCUSSION

Our finding of an intraclass correlation coefficient of 0.62 for admission scores on the original GAF agreed with Endicott et al.'s report⁴ of ICC's ranging from 0.61 to 0.91. Our ICC of 0.62 was significant at $P < 0.001$, thus indicating that while the reliability was somewhat low for admission ratings, it still was perfectly usable. Likewise, the ICC for discharge ratings from the original GAF was 0.90, which indicates excellent reliability. The value of the modified GAF (with its admission ICC of 0.81 and discharge ICC of 0.95) is for instances

when interrater reliability needs to be as high as it can be or when multiple persons of varying employment backgrounds and without much GAF training will rate patients. Research is a prime example for both uses of the modified GAF. Usually during research studies, there would also be enough time to read this longer GAF and assign ratings.

Another use for the modified GAF, compared with the original GAF or GAS, is in evaluating the need for hospital admission. Specifically, Thompson et al.,² in a review of 9,055 adult intakes, found marked variations in the way managed care case managers, compared with providers, assigned GAS scores generated from the same data. Thompson and colleagues felt that higher (less sick) scores reflected a need by managed care companies to limit the use of all inpatient services rather than their desire to selectively eliminate unnecessary hospitalizations. The ability of the managed care industry to affect the GAS scores in this way is attributed to the relatively less-structured nature of the GAS instrument, leading to lower interrater reliability. As we have shown, the modified GAF is both more structured than the original GAF or GAS and has better interrater reliability on admission scores. Thus, the modified GAF is less likely to reflect a bias by a managed care or governmental agency.

In addition to reliability tests, modified GAF ratings were also correlated with Zung depression tests and self-ratings of illness severity in outpatients. Similar to reliability tests, these correlations were in the same range as the correlations that Endicott et al.⁴ found between the original GAS and the Mental Status Examination Record (MSER) or the Family Evaluation Form (FEF) in outpatients. The slightly higher correlation between the modified GAF and the Zung depression test (-0.71), compared with the original GAS and MSER (0.62), probably was because all of our patients were depressed and the Zung specifically assessed depression. In contrast, the MSER is a global rating scale like the GAS, and there was probably greater heterogeneity among these patients. However, both of these sets of correla-

tions were interviewer information GAF each ingly, both that we cor the FEF, co GAS, gave the interviee the correlat self-rated s -0.45 for t Zung is al tions are global illn accounted with the C reported ti sician or may differ the same c

References

1. Western without chiatry
2. Thomps level of health 43:599-
3. McGlas chotrop of Men

tions were acceptable, thereby indicating that interviewer rated scales provide similar types of information and the original GAS and modified GAF each show acceptable validity. Interestingly, both the self-rated illness severity test that we correlated with the modified GAF and the FEF, correlated by Endicott with the original GAS, gave scores based on someone other than the interviewer's judgment, specifically the patient or the patient's family. Both of these sets of correlations were fairly low, 0.58 for the self-rated scale and modified GAF and -0.52 or -0.45 for the FEF and original GAS. While the Zung is also a self-rated instrument, its questions are more objective than the self-rated global illness scale or FEF, which may have accounted for the Zung's higher correlation with the GAF. Still, McGlashan⁵ and Pfeiffer⁶ reported that patient self-assessments and physician or interviewer assessments of patients may differ significantly. One might also expect the same discrepancy between family and inter-

viewer assessments of patients. Thus, the interviewer vs. self- or family-rating procedures for measuring severity of illness often cannot be considered as providing similar or redundant information.

The modified GAF is an instrument having a higher reliability and similar validity to the original GAF or GAS. The modified GAF may be particularly useful when interrater reliability needs to be maximum (i.e., in research or as a tool to determine need for hospitalization) and/or when multiple persons of varying skills and employment backgrounds and without having had much GAF training (i.e., in managed care organizations) must rate patients. In addition, when used to evaluate the need for hospital admission, the modified GAF is less likely than the original GAF or GAS to reflect a provider or managed care bias. Thus, our modified GAF may be a better and improved patient assessment tool, one that can more accurately reflect a patient's true need for hospitalization.

References

1. Westermeyer J: Problems with managed psychiatric care without a psychiatrist-manager. *Hosp Community Psychiatry* 1991; 42:1221-1224
2. Thompson JW, Burns BJ, Goldman HH, et al: Initial level of care and clinical status in a managed mental health program. *Hosp Community Psychiatry* 1992; 43:599-603
3. McGlashan T (ed): *The Documentation of Clinical Psychotropic Drug Trials*. Rockville, MD, National Institute of Mental Health, 1973
4. Endicott J, Spitzer RL, Fleiss JL, et al: The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Arch Gen Psychiatry* 1976; 33:766-771
5. McGlashan TH: The chestnut lodge follow-up study II. Long-term outcome of schizophrenia and the affective disorders. *Arch Gen Psychiatry* 1984; 41:586-601
6. Pfeiffer SI: An analysis of methodology in follow-up studies of adult inpatient psychiatric treatment. *Hosp Community Psychiatry* 1990; 41:1315-1321